Semantic Annotation and Retrieval using Multimedia Ontologies

Tomo Sjekavica, Ines Obradović, and Gordan Gledec

Abstract**—**Today's major problem in consumption of multimedia content from the Web is the extremely large volume of multimedia content in various forms on the Web, which keeps on rapidly growing. Another problem is that one part of that multimedia content is not annotated; therefore it is very hard to find and reuse such content. The other part of multimedia content is described manually, hence those annotations may be too subjective or inaccurate, and may be lacking in formal semantics. This results in the need for efficient semantic annotation, so that computers and applications can easily process those metadata for reuse and retrieval of multimedia content. This paper presents ontologies in general as part of Semantic Web and specific ontologies used for multimedia annotation. Comparison of the most commonly used multimedia ontologies and their main features is provided in this paper. These multimedia ontologies can be used for creating high quality and semantically rich multimedia annotations.

Keywords—metadata, multimedia ontologies, ontology, Ontology Design Patterns, OWL, semantic annotation, Semantic Web

I. Introduction

MULTIMEDIA content in all forms is every day taking more and more place in the web-available content. Most common types of multimedia content on the Web are images and video, but it can also be in form of 3D graphics, audio and audiovisual files. Besides of the consumption of multimedia content on the Web there is also a progressively increasing trend in amateur and professional production, which includes publishing that multimedia content on various User Generated Content (UGC) web sites, like Picasa, Flickr and YouTube [1]. Those sites do not enforce their users to make metadata definitions and to perform classification operations when uploading their multimedia content. With that large expansion of multimedia content on the Web, the need for indexing and annotating that content for efficiently use, reuse and retrieval of such content has occurred.

Multimedia content is annotated with metadata which adds additional value for that content. First type of multimedia metadata was plain text usually entered manually, which is time consuming and costly process. That kind of metadata is easily readable to humans, but computers can hardly process those metadata due to lack of formal semantics. Today a lot of different multimedia metadata standards and formats exist, like Exif, Dublin Core, VRA Core, DIG-35 and MPEG-7 that are not mutually compatible. MPEG-7 [2] is an international ISO/IEC multimedia content description standard that supports some degree of interpretation of information meaning, which can be processed by applications and computers, instead of just presenting information to the people. It is used for metadata of audiovisual content that can be in form of still pictures, graphics, 3D models, speech, audio or video. In order to enable better retrieval, discovery and exploitation of multimedia content on the Web by web services and applications there is a need for semantic annotation of multimedia content.

In order to achieve semantically rich annotations, the use of Semantic Web is required [3]. Semantic Web is an extension of the World Wide Web in which information is given well-defined meaning that enables better cooperation of computers and humans [4]. For semantic annotation of multimedia content Semantic Web technologies like XML, RDF and ontologies can be used. The common vocabulary representing shared knowledge within a specific domain can be defined with ontologies using final list of terms and concepts [5]. For humans, ontologies provide better access to information defined in ontology. Definitions of terms and concepts, as well as the relationships between them should enable better processing by applications and computers.

Although several vocabularies that can be used for semantic annotation of multimedia exist, they aren't rich enough or suitable for describing multimedia content for the use on the Semantic Web. Thus there is a need for development of extended, multimedia enriched ontologies, also known as multimedia ontologies.

This paper is organized as follows. Next section deals with ontologies in general and ontologies as part of the Semantic Web. An overview of ontology languages on the Web is provided in third section. Main Ontology Design Patterns are described in fourth section. Multimedia ontologies most commonly used for semantic annotation are shown in the fifth section. These selected multimedia ontologies are then

T. Sjekavica is with the Department of Electrical Engineering and Computing, University of Dubrovnik, Cira Carica 4, 20000 Dubrovnik, CROATIA (phone: +385 (0) 20 445 793; e-mail: tomo.sjekavica@unidu.hr).

I. Obradović is with the Department of Electrical Engineering and Computing, University of Dubrovnik, Cira Carica 4, 20000 Dubrovnik, CROATIA (e-mail: ines.obradovic@unidu.hr).

G. Gledec is with the University of Zagreb, Faculty of Electrical Engineering and Computing, Unska 3, 10000 Zagreb, CROATIA (e-mail: gordan.gledec@fer.hr).

compared in the sixth section. The seventh section, the last one before conclusion, provides an overview of related researches showing various approaches and methods for creating semantically rich annotation and semantic multimedia retrieval using different types of ontologies.

II. Ontologies

Term ontology originates from philosophy and has different meaning in different communities. There are various definitions of ontology in the literature. In more recent years term ontology was given specific technical meaning that is used in computer science. Instead of using „ontology“ the computer science uses „an ontology“.

A. Definitions of ontology

In computer science, Gruber defined ontology in 1993 as “an explicit specification of a conceptualization” [6]. Borst in his PhD thesis in 1997 modified Gruber’s definition, and provided new definition: “an ontology is a formal specification of a shared conceptualization” [7]. In 1998, Studer et al. combined above mentioned definitions and made a new definition for an ontology that is nowadays mostly used. Their definition states that “an ontology is a formal, explicit specification of a shared conceptualization” [8]. In this definition, a “conceptualization” is referring to an abstract model of some phenomena in the world, identifying the relevant concepts of those phenomena. “Formal” means that ontology should be machine readable. “Explicit” refers to the fact that type of used concepts and the constraints on their use must be explicitly defined. And finally “shared” reflects to the idea that an ontology is capturing a consensual knowledge that is accepted by group, instead of being private to some individual.

B. Classification of ontologies

There are different types of ontologies and they can be classified according to the object of conceptualization into four general levels [8] like shown in Fig. 1.

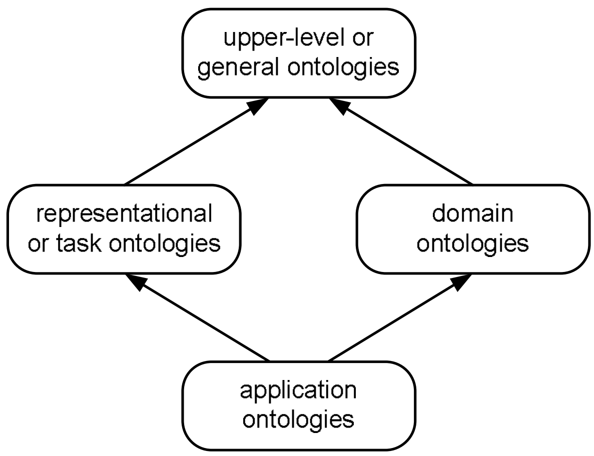

Fig. 1 Classification of ontologies according to the object of conceptualization

General or upper-level ontologies can be used across multiple domains describing very general concepts. Domain ontologies are more particular and they are used for a specific domain. Representational or task ontologies are not related to any specific domain and they provide representational entities without defining what they should represent. Application ontologies contain necessary knowledge for modeling a particular domain combining both task and domain ontologies.

C. Ontology hierarchy

Typically, an ontology is consisted of a finite list of classes (concepts) and the relationships between those classes. Relationships commonly include hierarchy of classes, which can specify that one class *Class1* can be subclass of another class *Class2* if every object in class *Class1* is included in the class *Class2*.

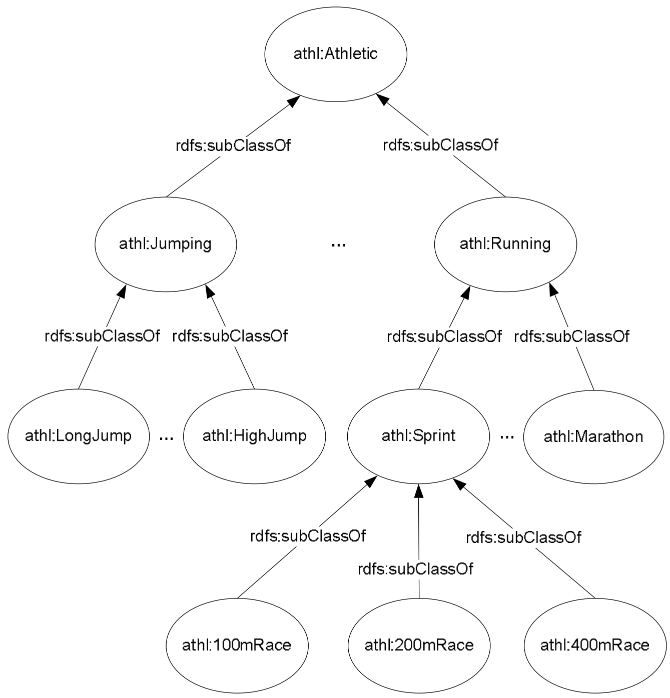

Fig. 2 Hierarchy of ontology classes in athletic domain

Fig. 2 shows an example for the hierarchy of ontology classes in athletic domain. In this example athletic domain can be divided into different categories of sports, like jumping and running. Furthermore those categories can then be divided into subcategories and/or disciplines. Thus running can be divided into subcategories from sprints to marathon races. In the sprint subcategory disciplines can be 100 meters, 200 meters and 400 meters races.

D. Ontology and the Semantic Web

Ontologies on the Web are usually used to enhance web search. Apart from the classic query string based web searches, ontologies provide effective semantic search of the web resources based on the context and content itself. Bonino et al. proposed a semantic search engine based on the ontology navigation in [9]. Ontology navigation provides semantic level reasoning for retrieving significant web resources based on given user queries. The key for a semantic web search is the availability of domain ontology and the ability for understanding the semantic relationships between concepts in ontology. Proposed semantic search engine uses semantic annotations about various web resources for providing relevant search results.

Reasons for developing ontologies and advantages of their use [5] are in:

- sharing common understanding of the structure of information among people and computers,
- analyzing and enabling reuse of domain knowledge,
- separating domain knowledge from the operational knowledge, and
- making explicit domain assumptions.

Tim Berners-Lee defined ontology as one of the main components of the Semantic Web and that most typical kind of ontology used on the Web has taxonomy and a set of inference rules [4]. On the Semantic Web, ontologies are commonly used in defining the meaning of resources and terms on the Web. Ontologies can also be used for semantic annotation and retrieval of multimedia content on the Web.

III. Ontology Languages

Users can write explicit and formal conceptualizations of domain models using different ontology languages. Main requirements [10] for every ontology language are:

- a well-defined syntax;
- efficient resoning support;
- a formal semantics;
- sufficient expressive power;
- convenience of expression.

Most commonly used language for ontologies on the Web today is Web Ontology Language (OWL). Predecessors of OWL are RDFS, SHOE, OIL, DAML-ONT and DAML+OIL. Major influence on the OWL design had DAML+OIL. On the other hand DAML+OIL was influenced by OIL language, with additional influence of RDFS and DAML-ONT.

A. DAML+OIL

The DARPA Agent Markup Language (DAML) was aimed at providing the foundation for the Semantic Web [11]. DAML would facilitate semantic interoperability with adoption of a common ontology language. Developed ontology language DAML-ONT extended RDF and its associated object-oriented type system. Along with development of DAML-ONT, Ontology Inference Language (OIL) which offers different levels of complexity was designed [12]. OIL combined elements from Description Logic (DL) [13] and frame-based systems along with XML and RDF elements.

DAML+OIL is a semantic markup language created as result of merging DAML-ONT and OIL [14]. It is designed to describe the structure of a domain. Formal semantics in DAML+OIL is given by its own DL style model theory instead of translation into a suitable DL [15]. It is integrated with RDF by the influence of DAML-ONT. DL constructors of OIL are used in DAML+OIL, while the frame structure was greatly discarded in a favor of DL style axioms that are more easily integrated with RDF syntax.

B. OWL

OWL [16] is a formal language used for publishing and sharing ontologies on the Semantic Web [15]. It is developed by W3C Web Ontology Working Group. Classes and properties can be defined using OWL, as well as relations between classes and characteristics of properties. A formal basis for the definition of the OWL was provided by DL. OWL is based on RDF and RDFS, and uses RDF/XML syntax. When the information needs to be processed by applications, instead of just presenting information to the people OWL is preferred to be used.

OWL provides three sublanguages with different levels of expressiveness [15] [16]:

- OWL Lite is used for simple ontologies with minimal expressiveness, where simple constraints and a classification hierarchy are considered of primary importance;
- OWL DL is based on description logic and it is used for expressive ontologies, where the maximum expressiveness is of primary importance, with the restrictions that all conclusions are guaranteed to be computable and that all computations will be completed in a finite time;
- OWL Full is syntactic and semantic extension of RDFS and it is used for maximum expressive ontologies where the compatibility with RDF and RDFS is of primary importance. It has the syntactic freedom, but does not give computational guarantees.

C. OWL 2

W3C OWL Working Group in 2009 created a new version of ontology language for the Semantic Web OWL 2 adding new features, while remaining compatible with the first version. OWL 2 has three profiles, known also as fragments or sublanguages, which are independent of each other [17]:

- OWL 2 EL can be used in applications which use ontologies with large number of properties and classes. The EL acronym refers that profile basis is in the EL family of DL that provide only Existential quantification;
- OWL 2 QL can be used where query answering is the most important reasoning task and in applications which use large volumes of instance data. The QL acronym refers to the fact that query answering can be

implemented by rewriting queries into standard relational Query Language;
- OWL 2 RL can be used in applications requiring scalable reasoning without sacrificing too much expressive power. The RL acronym refers to the fact that reasoning can be implemented using a standard Rule Language.

All OWL sublanguages provide additional formal vocabulary with added formal semantics allowing better communication with applications and greater machine interoperability of different content on the Web, than XML, RDF and RDFS provide. Multimedia ontologies created using OWL will enable creation of high quality and semantically rich multimedia metadata.

IV. Ontology Design Patterns

In the field of computer science, term "design pattern" stands for a general repeatable solution to a commonly occurring problem in software design. It is not a finished design but merely a description or template for how to solve a problem that can be used in many different situations. More recently, this notion has also appeared in requirements analysis, conceptual modelling, and ontology engineering [18]. Ontology design patterns (ODPs) are an emerging approach that favors the reuse of encoded experiences and good practices [19]. In ontology engineering, ODPs are used to facilitate or improve the techniques used during ontology lifecycle. According to Gangemi and Presutti [20], ODPs are modeling solutions to solve a recurrent ontology design problems. However, compared with software design patterns that have been used for a long period, patterns in ontology engineering are still at the very beginnings.

Here, we present some of the main ODPs used in development of ontologies for multimedia annotation.

A. Descriptions and Situations

Descriptions and Situations (D&S) was developed as an extension to the DOLCE foundational ontology in the EU WonderWeb project, presented in [21]. It formalizes the general paradigm of a situation and entities that constitute it (are the setting for it), and various means of describing this situation.

In general, D&S is an ontology of descriptions based on the widespread and very ancient ontological distinction between flux, or an unstructured world or context, and logos, or an intentionality. The D&S pattern provides an approach for context reification through a clear separation of „states-of-affairs" (i.e. set of assertions) and their interpretation based on a non-physical context, called a „description". D&S axioms try to capture the notion of "situation" as a unitarian entity out of a "state of affairs", where the unity criterion is provided by a "description". In that way, when a description is applied to a state of affairs, a situation emerges.

As Fig. 3 shows, the D&S pattern consists of a "Situation" that satisfies a "Description". The "Description" defines "Concepts". Each "Concept" classifies an "Entity". The entities are the individuals that are relevant in a given context. Each "Entity" is connected to the "Situation" by the *hasSetting* relation.

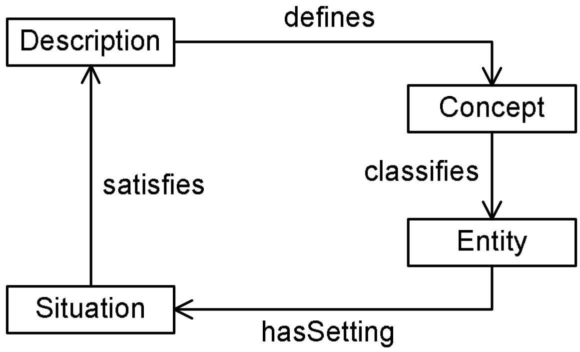

Fig. 3 Description and Situation pattern

D&S results to be a theory of ontological contexts because it is capable to describe various notions of context (physical and non-physical situations, topics, plans, assessments, beliefs, etc.) as first-order entities.

B. Ontology for Information Object

Ontology for Information Object (OIO) is adapted and improved from DOLCE foundational ontology [22], also developed within the EU WonderWeb project. Information Object (IO) presents type of social object and it is assumed to be equivalent to a content or information transferred in any modality. Fig. 4 shows basic IO design pattern. Properties of the basic IO design pattern are:

- *support* – realizes the IO;
- one or more *combinatorial structures* – IO is ordered according to those structures;
- *meaning* or *conceptualization* – IO expresses it;
- *reference* – what an IO is about;
- one or more *agents* – agents interpret the IO.

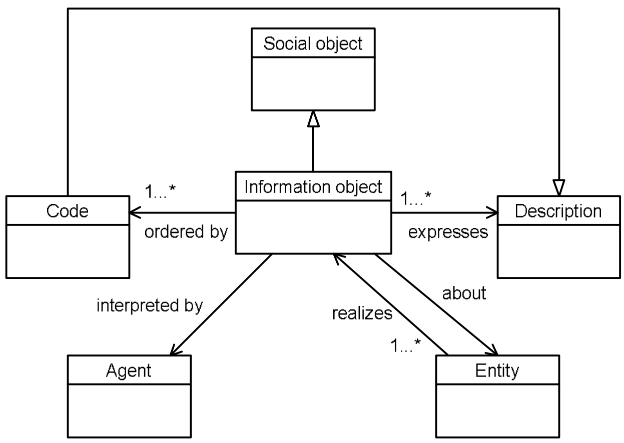

Fig. 4 Basic Information Object design pattern

IO design pattern can be used for modeling of domain specific information, like presented in [23]. Three main aspects differ in IO design pattern: “InformationRealization”, “InformationObject” and “Description”. The data or actual content are handled by “InformationRealization”. “Information Object” presents conceptualization, while “Description” provides information about data, content or subject matter. For example let’s take a digital image of beach sunset. “InformationRealization” will be that digital image, while the Description will be metadata of image. “InformationObject” will be concept of “Beach sunset”. In general “InformationObject” is *about* “InformationRealization” and *expresses* “Description”. Fig. 4 shows basic IO design pattern.

C. Information and Realization

The Information and Realization pattern [20] is a simple ODP that models the distinction between information objects such as poems, songs, formulas, and their physical realizations like printed books, registered tracks, physical files, etc. It is extracted from the DOLCE+DnS Ultralight (DUL) ontology as a special form of OIO pattern.

The pattern, as shown in Fig. 5, consists of the two main elements, which are the „InformationObject“ (a piece of information) and the „InformationRealization“ (a concrete realization of an „InformationObject“). A relation between an information realization and an information object is *realizes*, or *isRealizedBy* in the opposite direction.

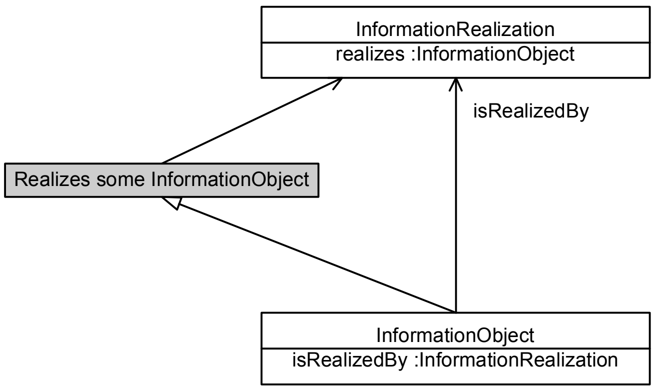

Fig. 5 Information Realization pattern

D. Data Value

With ontologies, abstract concepts and clearly identifiable individuals are used to represent data and to perform inferencing over the data. However, there is also the need for the means to represent concrete data values such as strings and numerical values [24]. Data Value pattern is also an ODP extracted from the DOLCE+DnS Ultralight ontology which is used to represent complex data values. In DUL, there is the concept “Quality” that represents attributes of an “Entity” (attributes that only exist together with the “Entity”). “Regions” are used to represent the values of “Quality” and the data space they come from.

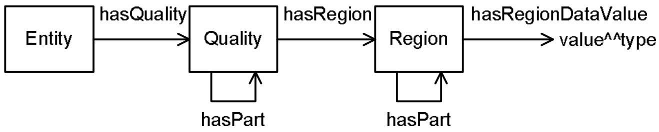

Fig. 6 Data Value pattern

The Data Value Pattern shown in Fig. 6 assigns a concrete data value to an attribute of that entity. The attribute is represented by the concept “Quality” and is connected to the “Entity” by the *hasQuality* property. The “Quality” is connected to a “Region” by the *hasRegion* relation. The “Region” models the data space the value comes from. The concrete value is attached to the “Region” using the relation *hasRegionDataValue*.

V. Multimedia Ontologies

Multimedia ontologies have been designed in order to serve one or more of the following tasks [25]:

- Annotation – tagging or labeling multimedia content;
- Analysis – ontology-driven semantic analysis of multimedia content;
- Retrieval – context-based image retrieval;
- Personalization – recommendation and filtering of multimedia content based on user preferences;
- Algorithms and processes control – modeling multimedia procedures and processes;
- Reasoning – personalization and retrieval for creating autonomous content applications.

In this section we provide an overview of the most common ontologies that are developed for use in multimedia domain and for annotation of multimedia content.

A. COMM

Arndt et al. in their work [26] address the need for a formal description of a high quality multimedia ontology. They compile and present the following list of requirements for a web-compliant multimedia ontology:

- MPEG-7 compliance,
- semantic interoperability,
- syntactic interoperability,
- separation of concerns,
- modularity and
- extensibility.

As a response, they propose the Core Ontology for Multimedia (COMM) which satisfies all of the requirements for a multimedia ontology framework. COMM (http://comm.semanticweb.org/) is an ontology implemented in OWL DL. The aim of COMM is to enable and facilitate multimedia annotation. It has been built re-engineering MPEG-7 standard and using DOLCE as its underlying foundational ontology. COMM is designed using two of the main ODPs: D&S and OIO, extending them for representation of MPEG-7 concepts.

The ontology covers a very large part of MPEG-7 standard. Moreover, COMM contains all MPEG-7 descriptors formalized using the same naming convention as in the MPEG-7 standard. The explicit representation of algorithms in the multimedia patterns allows also describing the multimedia analysis steps, something that is not possible in MPEG-7. The ontology is modularized to the core module and to modules specialized on each media type, i.e. the visual, text, media, localization and datatype modules, which minimizes execution overhead when processing data.

Additionally, for simplifying the creation of multimedia annotations, COMM provides a Java Application Programming Interface (API) which enables an MPEG-7 class interface for the construction of metadata at runtime.

B. Ontology for Media Resources 1.0

Ontology for Media Resources 1.0 (http://www.w3.org/TR/mediaont-10/) is both a core vocabulary (a set of properties describing media resources) and its mapping to a set of metadata formats currently describing media resources published on the Web [27]. It is developed by the W3C Media Annotations Working Group. The purpose of the mappings is to provide an interoperable set of metadata, thereby enabling different applications to share and reuse these metadata. The ontology targets at a unifying mapping of common media formats. An extensive set of mappings to many common multimedia metadata formats is provided. It recognizes 18 multimedia metadata formats (Dublin Core, MPEG-7, IPTC, Exif, OGG, etc.) and six multimedia container formats (3GP, FLV, QuickTime, MP4, OGG, WebM). The annotation properties include terms such as identifier, title, creator, date, location, description, keyword, rating, copyright, target audience, format, etc. The set of properties often has equivalence with existing formats. Therefore, a mapping table that defines one-way mappings between the ontology's properties and the metadata fields from other standards is specified. The proposed ontology has been formalized using an OWL representation. The ontology is also accompanied by an API that provides uniform access to all its elements.

C. M3O

Multimedia Metadata Ontology (M3O) is an ontology developed within the weKnowIt (http://www.weknowit.eu/) project for annotating rich, structured multimedia content on the Web and unlocking its semantics by making it machine-readable and machine-understandable. Saathoff and Scherp [28] proposed M3O in 2010 providing a generic modeling framework to integrate existing multimedia metadata formats and metadata standards. It bases on Semantic Web technologies and can be easily integrated with today's presentation formats like SMIL, SVG or Flash. The aim of M3O is integrating and representing the metadata and data structures that underlie the existing approaches, rather than replacing any of the existing models.

The M3O provides patterns that satisfy five principal requirements:

1. identification of resource,
2. separation between information objects and realizations,
3. annotation of information objects and information realizations,
4. decomposition of information objects and information realizations, and
5. representation of provenance information.

To meet these requirements, M3O represents data structures in form of patterns based on the foundational ontology DOLCE+DnS Ultralight. Three patterns specialized from DUL are reused in the M3O: D&S, Information and Realization, and Data Value. In addition, M3O provides Annotation and Decomposition patterns. Making use of these patterns, the ontology clearly distinguishes between the information object and its realization. It supports both the representation of high-level semantic annotation with background knowledge as well as the annotation with low-level features extracted from the multimedia content. The M3O has been aligned with COMM, Ontology for Media Resources and EXIF.

The M3O is available in OWL on the web (http://m3o.semantic-multimedia.org/). In addition, the API has been implemented [29] to make use of it in concrete multimedia applications.

D. LSCOM

Large-Scale Concept Ontology for Multimedia (LSCOM) [30] (http://vocab.linkeddata.es/lscom/) defines a formal vocabulary that includes more than 2.000 concepts for the annotation and retrieval of broadcast news video. The ontology was designed to satisfy multiple criteria of utility, coverage, feasibility, and observability. Concepts in LSCOM are related to the objects, activities and events, scenes and locations, people, programs, and graphics.

Under the LSCOM project, with the support of National Institute of Standards and Technology (NIST) and other US government agencies, on-going series of workshops TREC Video Retrieval Evaluation (TRECVID) are maintained (http://www-nlpir.nist.gov/projects/trecvid/). The goal of TRECVID workshops is to encourage researches in information retrieval by providing large test collection, uniform scoring procedures and forum for comparing results. In TRECVID 2012 workshop various research organizations completed one or more of six tasks on large scale test collection of video content from various sources [31]:

1. Semantic indexing (SIN),
2. Known-item search (KIS),
3. Instance search (INS),
4. Multimedia event detection (MED),

5. Multimedia event recounting (MER), and
6. Surveillance event detection (SER).

VI. Comparison of Selected Multimedia Ontologies

Table I shows basic information on the construction of selected multimedia ontologies. Table II shows the number of classes and object properties of the ontologies. Supported types of multimedia content by each of the multimedia ontologies are given in Table III.

TABLE I.
MULTIMEDIA ONTOLOGIES COMPARISON

Ontology	Language	Base ontology and multimedia metadata standard	Used Ontology Design Patterns
COMM	OWL DL	DOLCE, MPEG-7	D&S, OIO
Ontology for Media Resources 1.0	OWL DL	N/A, N/A	N/A
M3O	OWL Full	DUL, N/A	D&S, Information and Realization, Data Value
LSCOM	OWL	N/A, N/A	N/A

TABLE II.
NUMBER OF CLASSES AND OBJECT PROPERTIES IN MULTIMEDIA ONTOLOGIES

Ontology	Number of classes	Number of object properties
COMM	39	10
Ontology for Media Resources 1.0	14	56
M3O	126	129
LSCOM	2639	22

TABLE III.
SUPPORTED TYPES OF MULTIMEDIA CONTENT BY MULTIMEDIA ONTOLOGIES

Ontology	Supported types of multimedia content			
	Image	Video	Audio	Audio Visual
COMM	✓	✗	✓	✗
Ontology for Media Resources 1.0	+/-	✓	✓	+/-
M3O	✓	✓	✓	+/-
LSCOM	✗	✓	✗	✗

All presented ontologies are used for semantic annotation of one or more types of multimedia content. There are a number of general classes that are represented and used in almost all ontologies. COMM is one of the first ontologies developed for the multimedia annotation. COMM has modular design using upper-level ontology and ODPs, thus facilitating its extensibility and easy integration with other domain ontologies, which makes it most commonly used ontology for annotating multimedia content. The best feature of Ontology for Media Resources is its set of mappings with a great range of different multimedia metadata formats that can be used for annotating various multimedia content on the Web. LSCOM ontology is built-up to be used for annotating video content. Main advantages of LSCOM ontology among other ontologies that can be used for annotation of video content are ongoing TRECVID workshops that are held under the LSCOM project every year. A number of researchers from several research organizations are trying to resolve different tasks tied to semantic annotation on the large scale test collection of various video materials on those workshops. So this ontology keeps on extending and enriching every year.

VII. Related Work

A lot of research has been done in recent years on semantic annotation, search and retrieval of the multimedia content on the Web. Most researchers use the advantages of Semantic Web and ontologies in creating quality annotations of multimedia content for efficient processing by applications.

Ontology based approach for the creation and search of multimedia content is shown in [32]. Ontology editor Protégé-2000 is used for defining necessary ontologies for image annotation. On the images of the apes use case, two ontologies should be defined: domain-specific ontology and photo annotation ontology. Domain-specific ontology for mentioned use case is the animal domain which contains background knowledge and vocabulary that describes domain specific image features. Photo annotation ontology differ three viewpoints: i) subject matter feature, ii) photo feature and iii) medium feature. Subject matter feature connects photo annotation ontology with domain-specific ontology. Metadata containing information of when, how and why photo was taken is defined using photo feature. The way how the photo is stored, like photo resolution or image file format is determined with medium features.

The ontology based image retrieval approach is presented in [33]. This approach aims to standardize image descriptions and to allow better understanding of the semantic contents. The goal of semantic metadata is to describe user's retrieval queries and semantic content in images in order to enable efficient later use and reuse of those images. Semantic meanings of the images and search queries in this approach are described in XML files based on the concepts defined in shared ontology. Similarity between images and search queries in ontology based image retrieval approach is proposed to be determined in two basic steps: extraction of combined concept entities and similarity comparison between image and retrieval query. Semantic metadata is converted to a set of combined concept entities after the extraction of combined concepts. Semantic similarity comparison between images and retrieval queries is done using combined concept entities.

WordNet-based automatic image annotation system is presented in [34]. WordNet (http://wordnet.princeton.edu/) is a lexical database for English language, and it can be used as lexical upper-level ontology. Using WordNet hierarchical structure great image datasets are collected from seven independent image search engines. For every non-abstract noun from WordNet lexical database all the images provided by each search engine are automatically downloaded. After uniform and duplicate images are removed, using PageRank method wrong images for every noun are deleted and every word is covered with only top 100 images. At the end image dataset is used to train Support Vector Machine (SVM) classifiers and WordNet voting scheme for automatic image annotation.

An innovative approach to the formal description of low-level image features based on COMM ontology is presented in [35]. It makes use of the ontology's visual module as well as the core module. Vacura et al. in their paper use the image of French midfielder Zinedine Zidane during the football match as an example image. They deal with the dominant color descriptor as an element of the MPEG-7 standard. The dominant color descriptor specifies a set of dominant colors in an arbitrarily shaped region, which is a part of the image or the whole image. Two attributes are used: *ColorIndex*, which stands for the value specifying the index of the dominant color in the selected color space, and *Percentage*, that is the percentage of pixels that have the associated color value. It is presented how COMM can be used directly or through its associated Java API.

VIII. Conclusion

In recent years, along with the expansion of Web 2.0 and social networks, an extreme growth of multimedia content on the Web is registered. That multimedia content is mostly in the form of images and videos. To enable enhanced use, reuse and retrieval of multimedia content from the Web, that content needs to be annotated. Several multimedia metadata standards and a number of vocabularies commonly used for annotating multimedia content exist today. Semantic Web technologies, like RDF and ontologies, provide well-defined meaning for the multimedia content, enabling better processing of their annotations by computers and applications. Formal language OWL, along with its sublanguages, is used for defining ontologies on the Semantic Web.

This paper presented a brief overview of ontologies in general and selected specialized multimedia ontologies that can be used for semantically rich multimedia annotation. Through different methods and approaches using presented ontologies, the progress in semantic annotation of multimedia content is shown. Provided comparison of selected multimedia ontologies in this paper shows that COMM is most commonly used ontology for multimedia annotation because of its modular design enabling easy extensibility and integration with other domain ontologies. Ontology for Media Resources provides a great level of interoperabilty through its set of mappings with a number of different multimedia metadata formats. LSCOM ontology is preferred to be used in annotation of video content, mainly because it keeps on extending every year through ongoing TRECVID workshops on which different tasks related to the semantic annotation and retrieval on the large scale test collection of video materials are resolved.

Our ongoing research is directed towards development of a new multimedia ontology based on one or more of the existing ontologies as its base underlying ontology. That new multimedia ontology should enable high quality and semantically rich multimedia annotations of images and photos.

References

[1] M. I. Lunescu, G. Concas, "Multimedia standard in UGC", *Proceedings of the 2nd WSEAS International Conference on Environment, Economics, Energy, Devices, Systems, Communications, Computers, Mathematics (EDSCM '12)*, Saint Malo & Mont Saint-Michel, France, 2012, pp. 242-246.

[2] J. M. Martínez, "MPEG-7 overview (version 10)", ISO/IEC JTC1/SC29/WG11, ISO/IEC, October 2004.

[3] T. Sjekavica, G. Gledec, M. Horvat, "Multimedia annotation using Semantic Web technologies", *Proceedings of the 7th WSEAS European*

Computing Conference (ECC '13), Dubrovnik, Croatia, 2013, pp. 228-233.

[4] T. Berners-Lee, J. Hendler, O. Lassila, "The Semantic Web", *Scientific American*, Vol. 284, No. 5, May 2001, pp. 34-43.

[5] N. F. Foy, D. L. McGuinness, "Ontology development 101: A guide to creating your first ontology", Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880, 2001.

[6] T. R. Gruber, "A translation approach to portable ontology specifications", *Knowledge Acquisition*, Vol. 5, Issue 2, 1993, pp. 199-220.

[7] W. N. Borst, *Construction of engineering ontologies for knowledge sharing and reuse* , PhD Thesis, University of Twente, Enschede, Netherlands, 1997.

[8] R. Studer, V. R. Benjamins, D. Fensel, "Knowledge engineering: principles and methods", *Data & Knowledge Engineering*, Vol. 25, No. 1-2, 1998, pp.161-197.

[9] D. Bonino, F. Corno, L. Farinetti, A. Bosca, "Ontology driven semantic search", *WSEAS Transaction on Information Science and Application*, Vol. 1, Issue 6, 2004, pp. 1597-1605.

[10] G. Antoniou, F. van Harmelen, *A Semantic Web Primer*, 2nd ed., The MIT Press, Cambridge, Massachusetts, USA, 2008.

[11] J. Hendler, D. L. McGuinness, "The DARPA Agent Markup Language", *IEEE Intelligent Systems*, Vol. 15, No. 6, 2000, pp. 67-73.

[12] D. Fensel, F. van Harmelen, I. Horrocks, D. L. McGuinness, P. F. Patel-Schneider, "OIL: an ontology infrastructure for the Semantic Web", *IEEE Intelligent Systems*, Vol. 16, No. 2, 2001, pp. 38-45.

[13] F. Baader, D. Calvanese, D. L. McGuinness, D, Nardi, P. F. Patel-Schneider, *The Description Logic handbook: Theory, implementation and applications*, 2nd ed., Cambridge University Press, 2010.

[14] D. Connolly, F. van Harmelen, I. Horrocks, D. L. McGuinness, P. F. Patel-Schneider, L. A. Stein, "DAML+OIL (March 2001) reference description", *W3C Note*, December 2001. [Online]. Available: http://www.w3.org/TR/daml+oil-reference/

[15] I. Horrocks, P. F. Patel-Schneider, F. van Harmelen, "From SHIQ and RDF to OWL: The making of a Web Ontology Language", *Journal of Web Semantics: Science, Services and Agents on the World Wide Web*, Vol. 1, No. 1, 2003, pp. 7-26.

[16] D. L. McGuinness, F. van Harmelen, "OWL Web Ontology Language overview", *W3C Recommendation*, February 2004. [Online]. Available: http://www.w3.org/TR/owl-features/

[17] B. Motik, B. C. Grau, I. Horrocks, Z. Wu, A. Fokoue, C. Lutz, "OWL 2 Web Ontology Language profiles (second edition)", *W3C Recommendation*, December 2012. [Online]. Available: http://www.w3.org/TR/owl2-profiles/

[18] A. Gangemi, "Ontology Design Patterns for Semantic Web Content", *Proceedings of the 4th International Semantic Web Conference (ISWC 2005)*, Galway, Ireland, 2005, pp.262-275.

[19] R. de Almeida Falbo, M. P. Barcellos, J. C. Nardi, G. Guizzardi, "Organizing Ontology Design Patterns as Ontology Pattern Languages", *ESWC*, Springer, 2013, pp. 61-75.

[20] A. Gangemi, V. Presutti, "Ontology Design Patterns", *Handbook on Ontologies*, 2nd ed., Springer, Berlin, 2009, pp. 221-244.

[21] A. Gangemi, P. Mika, "Understanding the Semantic Web through Descriptions and Situations", *Proceedings of ODBASE03 Conference*, Springer, 2003, pp. 689-706.

[22] A. Gangemi, S. Borgo, C. Catenacci, J. Lehmann, "Task taxonomies for knowledge content D07", Tehcnical report, Metokis Deliverable 7 v.2.1., 2005.

[23] N. Arora, R. Westenthaler, W. Behrendt, A. Gangemi, "Information object design pattern for modeling domain specific knowledge", *Proceedings of the 1st ECOOP Workshop on Domain-Specific Program Development (DSPD) in conjunction with ECOOP*, Nantes, France, 2006.

[24] D. Eißing, A. Scherp, C. Saathoff. "Integration of existing multimedia metadata formats and metadata standards in the M3O*", Proceedings of the 5th international conference on Semantic and digital media technologies*, SAMT'10, Berlin, Germany, 2011, pp. 48–63.

[25] Y. Kompatsiaris, P. Hobson, *Semantic multimedia and ontologies: theory and applications*, Springer, London, 2008.

[26] R. Arndt, R.Troncy, S. Staab, L. Hardman, "COMM: A core ontology for multimedia annotation", *Handbook on Ontologies*, 2nd ed., Springer, Berlin, 2009, pp. 403-421.

[27] W. Lee, W. Bailer, T. Bürger, P.A. Champin, J.P. Evain, V. Malaisé, T. Michel, F. Sasaki, J. Söderberg, F. Stegmaier, J. Strassner, "Ontology for Media Resources 1.0"**,** *W3C Recommendation*, February 2012. [Online]. Available: http://www.w3.org/TR/mediaont-10/

[28] C. Saathoff, A. Scherp, "Unlocking the semantics of multimedia presentations in the web with the multimedia metadata ontology", *WWW '10 Proceedings of the 19th international conference on World Wide Web*, 2010, pp. 831-840.

[29] A. Scherp, C. Saathoff, "A pattern system for describing the semantics of structured multimedia documents", *International Journal of Semantic Computing 6(3),* 2012, pp. 263-288.

[30] M. Naphade, J. R. Smith, J. Tesic, S.-F. Chang, W. Hsu, L. Kennedy, A. Hauptmann, J. Curtis, "Large-Scale Concept Ontology for Multimedia", *IEEE MultiMedia*, Vol. 13, No. 3, 2006, pp. 86-91.

[31] P. Over, J. Fiscus, G. Sanders, B. Shaw, M. Michel, G. Awad, A. F. Smeaton, W. Kraaij, G. Quénot, "TRECVID 2012 – An overview of the goals, tasks, data, evaluation mechanisms, and metrics", *Proceedings of TRECVID 2012*, NIST, USA, 2012.

[32] A. T. Schreiber, B. Dubbeldam, J. Wielemaker, B. Wielinga, "Ontology-based photo annotation", *IEEE Intelligent Systems*, Vol. 16, No. 3, 2001, pp. 66-74.

[33] W. Zheng, Y. Ouyang, J. Ford, F. S. Makedon, "Ontology-based image retrieval", *Proceedings of 5th WSEAS International Conference on Mathematical Methods and Computational Techniques in Electrical Engineering (MMACTEE '03)*, Vouliagmeni, Athens, Greece, 2003.

[34] J. Lu, Z. Lu, Y. Li, T. Zhao, Y. Zhang, "A new large-scale image automatic annotation system based on WordNet", *Education Technology and Training, 2008 and 2008 International Workshop on Geoscience and Remote Sensing. ETT and GRS 2008. International Workshop on*, Vol. 1, 2008, pp. 758-762.

[35] M.Vacura, V.Svátek, C.Saathoff, T. Ranz, R. Troncy, "Describing low-level image features using the COMM ontology", *Proceedings of the 15th International Conference on Image Processing (ICIP)*, San Diego, 2008, pp. 49-52.